

СПЕЦИФИЧЕСКИЕ ПРЕФИКСОИДЫ В КОРПУСЕ РУССКОЯЗЫЧНЫХ СМИ ГРОДНЕНЩИНЫ

Корпус русскоязычных СМИ Гродненщины (далее — РМ(Г)) создан в рамках российско-белорусского проекта, направленного на изучение лексико-семантической и социокультурной специфики русской речи вне метрополии; ресурс включен в Корпус региональной и зарубежной прессы НКРЯ [1].

Согласно техническому заданию проекта, выявление языковых единиц, специфических для РМ(Г), осуществлялось с обращением к корпусу-эталону (языковому ресурсу, сопоставимому с РМ(Г) и отражающему узус русской речи в региональных СМИ России).

Предоставленный российской стороной корпус-эталон российских СМИ (далее — КЭ) содержит тексты изданий центральных и северо-западных областей России, юга России, Поволжья, Урала, Сибири, Дальнего Востока; количественный состав корпуса-эталона (без учета регистра): 176 977 словоформ (2 301 308 словоупотреблений), реализованных в 3 681 тексте. Поскольку РМ(Г) представляет собой языковой ресурс смешанного типа, в котором выделяются русскоязычная, смешанная и белорусскоязычная части, для корректности процедуры сопоставления с КЭ поиск специфических префиксоидов проводился на массиве русскоязычной и смешанно-русскоязычной частей РМ(Г) (далее — РМ(Г_{см})); количественный состав РМ(Г_{см}) (без учета регистра): 115 414 словоформ (1 261 088 словоупотреблений), реализованных в 3 862 тексте.

Префиксоид понимаем как корневую морфему, которая регулярно сочетается как минимум с двумя основами, соотносимыми с разными корнями, и демонстрирует при этом устойчивую тенденцию существенного расширения круга таких основ. В этой работе рассмотрены префиксоиды, дополнительным формальным признаком которых является наличие конечного дефиса и (опционально) варианта записи без дефиса на массиве РМ(Г_{см}) и/или КЭ (например: *арт-* / *арт:* *арт-хаус*, *арт-проект* и *артхаус*, *артпроект*; *стоп-* / *стоп:* *стоп-кран*, *стоп-линия*, *стоп-мен* и *стопмен*; *пожарно-*: *пожарно-аварийный*, *пожарно-прикладной* и др.).

По признаку «несклоняемость первой части» слова с такими префиксоидами отделены от сложных слов, созданных по модели родовое-видовое и восполняющих номинативную лакуну: *студент*-{*бюджетник, платник, иностранец, историк, минчанин, хорист, юрист*}, *врач*-{*гериатр, гигиенист, иммунолог, косметолог, рентгенолог, эпидемиолог, эпизоотолог*}, *ярмарка*-{*продажа, вакансия*} и т.п.

Префиксоиды выделялись следующим образом:

- из частотного словника $PM(\Gamma_{\text{м}})$ в автоматизированном режиме были выделены неединичные на $PM(\Gamma_{\text{м}})$ конструкты формата «набор букв + дефис»; неединичность понималась как одновременное соответствие двум условиям:

- а) частота конструкта в $PM(\Gamma_{\text{м}})$ выше 1;
- б) конструкт сочетается с разными основами;

- результат был сохранен в формате: «конструкт; частота конструкта в $PM(\Gamma_{\text{м}})$; частота конструкта в КЭ»; если конструкт отсутствовал в КЭ, назначалась соответствующая частота 0;

- из полученного массива записей вручную были исключены нецелевые единицы;

- если у префиксоида был реализован вариант без конечного дефиса, соответствующая запись редактировалась (в том числе и по параметру частоты).

Для оценки специфичности префиксоида, выделенного на массиве $PM(\Gamma_{\text{м}})$, вычислялось значение меры неслучайности (метод вычисления предложен в работе А.Я. Шайкевича и С.О. Савчук [2]).

Мера неслучайности S вычислялась по формуле:

$$S = (f - m - 1) / \sqrt{m}$$

где f — частота префиксоида в $PM(\Gamma_{\text{м}})$, m — математическое ожидание частоты префиксоида в $PM(\Gamma_{\text{м}})$ как части $PM(\Gamma_{\text{м}}) + \text{КЭ}$.

Приведем список выделенных префиксоидов со значимым для выделения специфических единиц значением $S \geq 2$:

молочно- / *молочно* ($S = 10$), *белорусско-* / *белорусско* ($S = 8$), *фото-* / *фото* ($S = 7$), *физкультурно-* ($S = 7$), *профессионально-* ($S = 7$), *коррекционно-* ($S = 6$), *немецко-* / *немецко* ($S = 6$), *туристско-* ($S = 6$), *учебно-* ($S = 6$), *эко-* / *эко* ($S = 6$), *оперативно-* ($S = 6$), *дорожно-* ($S = 5$), *осенне-* ($S = 4$), *весенне-* ($S = 4$), *семейно-* ($S = 4$), *духовно-* ($S = 4$), *историко-* ($S = 4$), *врачебно-* ($S = 4$), *справочно-* ($S = 3$), *товарно-* ($S = 3$), *мини-* ($S = 3$), *культурно-* ($S = 3$), *арт-* / *арт* ($S = 3$), *sos-* ($S = 3$), *свето-* ($S = 3$), *трофи-* ($S = 3$), *мясо-* / *мясо* ($S = 3$), *дворцово-* ($S = 3$), *комплексно-* ($S = 3$), *gps-* ($S = 2$), *стоп-* / *стоп* ($S = 2$), *ретро-* / *ретро* ($S = 2$), *административно-* ($S = 2$), *аграрно-* ($S = 2$), *конно-* / *конно* ($S = 2$), *польско-* / *польско* ($S = 2$), *сердечно-* ($S = 2$), *чудо-* / *чудо* ($S = 2$).

Критерий S позволяет выделить специфически высокочастотные префиксоиды, но не учитывает их деривационной активности: например, одинаковое значение $S = 7$ присваивается префиксоидам *физкультурно-* и *фото-* / *фото*, очевидно различным по деривационному потенциалу: пре-

фиксоид *физкультурно-* в каждом из корпусов сочетается с двумя элементами: *оздоровительный, спортивный*; существенно более деривационно активен префиксоид *фото-* / *фото:* в РМ($\Gamma_{\text{м}}$) сочетается с 49 основами, 32 из которых не дублируются в КЭ, этот же префиксоид в КЭ сочетается с 41 основой, 24 из которых не дублируются в РМ($\Gamma_{\text{м}}$).

В качестве рабочего критерия оценки деривационной активности префиксоидов вводим коэффициент D :

$$D = \sqrt{U_1 * IPM_1} - \sqrt{U_2 * IPM_2},$$

где U_1 и U_2 — число уникальных основ, сочетающихся с префиксоидом в РМ($\Gamma_{\text{м}}$) и в КЭ соответственно, IPM_1 и IPM_2 — значения ИРМ группы слов с таким префиксоидом в РМ($\Gamma_{\text{м}}$) и в КЭ соответственно.

Ниже приведен список префиксоидов РМ($\Gamma_{\text{м}}$) со значением $D \geq 0$: *фото-* / *фото* ($D = 31$), *белорусско-* / *белорусско* ($D = 21$), *эко-* / *эко* ($D = 20$), *молочно-* / *молочно* ($D = 14$), *арт₁-* / *арт₁* ($D = 12$), *профессионально-* ($D = 11$), *туристско-* ($D = 11$), *оперативно-* ($D = 10$), *коррекционно-* ($D = 9$), *семейно-* ($D = 7$), *немецко-* / *немецко* ($D = 7$), *медико-* ($D = 6$), *дорожно-* ($D = 6$), *справочно-* ($D = 6$), *свето-* ($D = 5$), *товарно-* ($D = 5$), *польско-* / *польско* ($D = 5$), *стоп-* / *стоп* ($D = 5$), *культурно-* ($D = 5$), *учебно-* ($D = 4$), *ретро-* / *ретро* ($D = 4$), *врачебно-* ($D = 4$), *сос-* ($D = 4$), *спа-* ($D = 4$), *трофи-* ($D = 4$), *конно-* / *конно* ($D = 4$), *пожарно-* ($D = 3$), *духовно-* ($D = 3$), *гprs-* ($D = 3$), *рентген-* / *рентген* ($D = 3$), *аграрно-* ($D = 3$), *литовско-* / *литовско* ($D = 3$), *онлайн-* ($D = 3$), *греко-* ($D = 3$), *бежево-* ($D = 3$), *ралли-* ($D = 3$), *гала* ($D = 3$), *веб-* ($D = 2$), *тест-* ($D = 2$), *дизель-* ($D = 2$), *линейно-* ($D = 2$), *допинг-* ($D = 2$), *лор-* ($D = 2$), *режимно-* ($D = 2$), *фан-* ($D = 2$), *роль-* ($D = 2$), *царь-* ($D = 2$), *online-* ($D = 2$), *рабоче-* ($D = 2$), *биолого-* ($D = 2$), *караоке-* ($D = 2$), *костно-* ($D = 2$), *чудо-* / *чудо* ($D = 1$), *мясо-* / *мясо* ($D = 1$), *фитнес* ($D = 1$), *розово-* / *розово* ($D = 1$), *блиц-* / *блиц* ($D = 1$), *танц-* / *танц* ($D = 1$).

Комментарии к списку.

1. Помимо префиксоида *арт₁-* / *арт₁* (*арт* — ‘искусство’), в РМ($\Gamma_{\text{м}}$) реализован префиксоид *арт₂-* / *арт₂* (*арт* — ‘артиллерийский’), деривационно активный в КЭ и префиксоид *арт₃-* / *арт₃* (*арт* — ‘артезианский’), не маркированный по активности в обоих корпусах.

2. Префиксоид *чудо-* / *чудо* ($D = 1$) представляет группу оценочных префиксоидов. Неспецифический для РМ($\Gamma_{\text{м}}$) префиксоид этой группы *горе-* / *горе* ($D = -1$) деривационно активен в КЭ.

3. Лексикографирование аффиксоидов может стать целью отдельной разработки. В приложениях к «Обратному словарю русского языка» [3] приведены списки таких единиц с примерами и пометами фиксации этих примеров в толковых словарях. Списки регионально специфических префиксоидов и связанных с ними основ могут рассматриваться как своеобразные справочники сочетаемости деривационно активных для Гродненского региона единиц; такой ресурс может быть использован в практике преподавания РКИ. В завершение приведем пример входа такого списка.

Фото- / фото (D = 31).

А. Уникальные для РМ($\Gamma_{\text{м}}$) сочетания с основами (32):

ФОТО-: загадка (3), карта (1), услуга (2).

ФОТО: ателье (2), вспышка (2), галерея (1), дебют (1), дело (1), динамический (1), издание (1), иллюстрация (1), искусство (2), клуб (11), книга (1), кор (2), мастер (1), омоложение (1), отчет (2), панорама (2), пейзаж (1), пленэр (1), презентация (1), радар (1), рассказ (1), студия (1), факт (3), фиксация (2), фрагмент (1), фьюжн (1), центр (1), школа (1), экспрессия (1).

Б. Уникальные для КЭ сочетания с основами (24):

ФОТО-: видеоохота (1), ню (1), сервис-центр (1);

ФОТО: архив (4), век (1), город (1), диод (1), журналист (4), затвор (1), информация (1), копия (1), лаборатория (2), любитель (1), модель (5), модельный (1), очерк (2), пластинка (1), проект (1), реализм (1), репортер (1), робот (1), стенд (1), телетехника (1), элемент (1).

С. Общие для РМ($\Gamma_{\text{м}}$) и КЭ сочетания с основами (17) :

ФОТО: альбом (8; 4), аппарат (9; 11), выставка (9; 4), камера (9; 4), конкурс (5; 2), корреспондент (6; 1), материал (1; 1), монтаж (1; 2), пленка (1; 2), портрет (2; 1), работа (9; 1), репортаж (2; 2), сессия (3; 3), синтез (1; 1), снимок (13; 2), съемка (3; 4), художник (8; 1).

Подготовлено в рамках проекта, реализуемого при поддержке БРФФИ (договор № Г13Р-050 от 16.04.2013).

ЛИТЕРАТУРА

1. Корпус региональной и зарубежной прессы / НКРЯ. [Электронный ресурс]. — Режим доступа: <http://www.ruscorpora.ru/search-regional.html>. — Дата доступа: 10.03.2015.

2. Шайкевич, А.Я., Савчук, С.О. Анализ лексико-семантических особенностей региональной прессы (на примере газет Гродненского региона Беларуси) / Компьютерная лингвистика и интеллектуальные технологии. Вып. 13 (20). [Электронный ресурс]. — Режим доступа: http://www.dialog-21.ru/digests/dialog2014/materials/pdf/ShaikevichAY_SavchukSO.pdf. — Дата доступа: 10.03.2015.

3. Обратный словарь русского языка. Ок. 125 000 слов. — М., 1974. — С. 894–906.